



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12Q 1/68, G01N 33/00, 33/48	A1	(11) International Publication Number: WO 92/05285 (43) International Publication Date: 2 April 1992 (02.04.92)
(21) International Application Number: PCT/US91/06793 (22) International Filing Date: 18 September 1991 (18.09.91) (30) Priority data: 586,769 21 September 1990 (21.09.90) US (71) Applicant: FRED HUTCHINSON CANCER RESEARCH CENTER [US/US]; 1124 Columbia Street, Seattle, WA 98104 (US). (72) Inventors: WEINTRAUB, Harold, Martin; 1223 16th Avenue East, Seattle, WA 98112 (US). BLACKWELL, Thomas, Keith; 333 Summit Avenue East, #403, Seattle, WA 98102 (US). (74) Agents: MURASHIGE, Kate, H. et al.; Morrison & Foerster, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).	(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, SE (European patent), SU*. Published With international search report.	
(54) Title: PROTEIN SEQUENCE-SPECIFIC OLIGONUCLEOTIDE SEQUENCES (57) Abstract A method for identifying oligomer sequences which specifically bind target proteins or other target moieties is described. The technique involves complexation of the target substance with a mixture of oligonucleotides containing random sequences and sequences which serve as primers for PCR under conditions wherein a complex is formed with some but not all members of the oligonucleotide mixture. The complex is then separated from uncomplexed oligonucleotides and the complexed members of the oligonucleotide mixture are recovered and amplified from the separated complex using the polymerase chain reaction. Successive rounds of selection using complexation, separation, amplification and recovery can be employed. The recovered mixture or oligonucleotide may be sequenced and this information used for the <i>de novo</i> synthesis of the identified oligonucleotide(s). These synthesized oligonucleotides may contain one or more modified bases or backbone linkages.		

* See back of page

INTERNATIONAL SEARCH REPORT

International Application No PCT/US91/06793

I. CLASSIFICATION OF SUBJECT MATTER (In several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC(5): C12Q 1/68; G01N 33/00, 33/48		
U.S.C1: 435/6; 436/63, 94, 501		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
U.S.C1:	435/6; 436/63, 94, 501	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No ¹³
Y	Nucleic Acids Research, volume 17 Number 10, issued 1989, Kenneth W. Kinzler et al., "Whole Genome PCR: Application to the Identification of Sequences Bound by Gene Regulatory Proteins," pages 3645-3653, see abstract, Figure 1, and page 3645, last paragraph-page 3650, line 6.	1-13
Y	BioTechniques, Volume 6, Number 10, issued 1988, vander Krol et al., "Modulation of Eukaryotic Gene Expression by complementary RNA or DNA Sequences," pages 958-976, see page 956, first paragraph.	11
<p>¹⁰ Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
01 November 1991	25 NOV 1991	
International Searching Authority	Signature of Authorized Officer	
ISA/US	Lori Yuan	

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU*	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TC	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

5 PROTEIN SEQUENCE-SPECIFIC OLIGONUCLEOTIDE SEQUENCESTechnical Field

10 The invention is directed to a method to identify oligonucleotide sequences which specifically bind target proteins. More specifically, it concerns a method to identify the appropriate oligonucleotide sequence for such binding, and several oligonucleotide sequences which correspond to proteins known to be instrumental in differentiation.

15

Background and Related Art

20 The scope of what was originally designated "antisense" therapy and diagnosis has expanded greatly in the last several years. The original concept sought to take advantage of the specific hybridization of DNA and RNA oligonucleotides to their complements to inactivate such specific DNA or RNA oligonucleotides which mediate diseases or other undesirable conditions in humans, animals, and even plants.

25 The origin of the term "antisense" is thus clear: the therapeutic or diagnostic oligonucleotide would be the antisense counterpart of the targeted RNA or DNA. The "antisense" oligonucleotides can be supplied directly or generated in situ and may either be
30 conventional oligomers, or are, more commonly, oligomers having properties which make them, for example, resistant to nucleases, more capable of transfer across membranes, or more capable of specific binding to the desired target. However, in addition to the specific binding
35 effected by conventional base pairing, the oligonucleo-

-2-

tides used in this approach may recognize double-stranded DNA by binding to the major or minor grooves present in the double-helix.

Such approaches have been suggested, for example, to interfere with transcription by binding to promoter sequences in duplexed DNA to prevent expression of the related gene. Therefore, the concept has expanded beyond a simple "antisense" approach to include any therapy by administration or in situ generation of oligonucleotides. The general approach to constructing various oligomers useful in "antisense" therapy has been reviewed by Van der Krol, A.R. et al., Biotechniques (1988) 6:958-976 and by Stein, C.A. et al., Cancer Res (1988) 48:2659-2668, both incorporated herein by reference.

The extension of oligonucleotide-based therapy to include binding to duplexed DNA was made possible by elucidation of the rules governing sequence-specific binding in this context. While not so precisely understood as the requirements for base-pair complementation, these principles have been sufficiently described to make de novo design of oligomers which will bind to known target duplexes possible. Such de novo design of specifically binding oligonucleotides is not, however, possible with respect to non-oligonucleotide targets. Formulation of an approach that would permit construction of oligonucleotides capable of specific binding to any desired target substance would clearly be desirable. By use of such oligonucleotides, the modulation of the metabolic events associated with any condition, disease, or developmental process for which any critical substance is known could be effected. Furthermore, the specifically binding oligonucleotides are useful in diagnostic and assay methods and in regulation of cell cultures in vitro. The method of the invention permits

-3-

just such design of oligonucleotides comprising sequences specific for any target substance of sufficient size to show complexation with DNA or RNA sequences.

The invention method utilizes the polymerase chain reaction (PCR) technique, as described by Saiki, R.K., et al., Science (1988) 239:487-491. There are a number of related publications which describe the use of this technique in similar contexts. For example, Joyce, G.F., Gene (1989) 82:83-87 applied the PCR reaction to plus strand RNA/minus strand DNA complexes to study the evolution of RNAs with catalytic activity. Various strategies for producing mutations in RNA to provide the catalytic activity are discussed. Robertson, D.L., and Joyce, G.F., in a letter to Nature (1990) 344:467-468, describe the results of application of this technique to obtain a catalytic RNA which cleaves DNA more efficiently than the wild-type enzyme.

Kinzler, K.W., et al., Nucleic Acids Res (1989) 17:3645-3653, applied this technique to identify DNA sequences that bind to proteins that regulate gene expression. In the reported work, total genomic DNA is first converted to a form that is suitable for amplification by PCR and the DNA sequences of interest are selected by binding to the target regulatory protein. The recovered bound sequences are then amplified by PCR. The selection and amplification process are repeated as needed. The process as described was applied to identify DNA sequences which bind to the Xenopus laevis transcription factor 3A. The same authors (Kinzler et al.) in a later paper, Mol Cell Biol (1990) 10:634-642, applied this same technique to identify the portion of the human genome which binds to the GLI gene product produced as a recombinant fusion protein. The GLI gene is amplified in a subset of human tumors.

35

-4-

Ellington, A.D., et al., Nature (1990) 346:818-822 describe the production of a large number of random sequence RNA molecules and identification of those which bind specifically to small ligands, in the case of this paper, to specific dyes such as Cibacron blue. Randomly synthesized DNA yielding approximately 10^{15} individual sequences was amplified by PCR and transcribed into RNA. It was thought that the complexity of the pool was reduced in the amplification/transcription steps to approximately 10^{13} different sequences. The pool was then applied to an affinity column containing the dye and the bound sequences subsequently eluted, treated with reverse transcriptase and amplified by PCR. The results showed that about one in 10^{10} random sequence RNA molecules folds in such a way as to bind specifically to the ligand.

Tuerk, C. and Gold, L. in Science (1990) 249:505-510 used what they referred to as the procedure of "systematic evolution of ligands by exponential enrichment" (Selex) which is described as follows: a pool of RNAs that are completely randomized at specific positions is subjected to selection for binding to a desired protein which has been displayed on a nitrocellulose filter. The selected RNAs are then amplified as double-stranded DNA that is competent for subsequent in vitro transcription. The newly transcribed RNA is then enriched for better binding sequences and recycled through this procedure. The amplified selected sequences are subjected to sequence determination using dideoxy sequencing. Tuerk and Gold applied this procedure to determination of RNA ligands which bind to T4 DNA polymerase.

Thiesen, H.-J. and Bach, C. Nucleic Acids Res (1990) 18:3203-3208 described what they call a target detection assay (TDA) to determine DNA binding sites for

-5-

putative DNA binding proteins. In their approach, a purified functionally active DNA binding protein and a pool of random double-stranded oligonucleotides which contain PCR primer sites at each end were incubated with the protein. The resulting DNA complexes with the protein (in their case, the SP1 regulatory protein) were separated from the unbound oligomers in the random mixture by band-shift electrophoresis and the complex oligonucleotides were rescued by PCR and cloned, and then sequenced using double-stranded mini-prep DNA sequencing.

The invention herein utilizes a binding site selection technique which also depends on the availability of PCR. In this approach, selected and amplified binding sites (SaABs) provide a characteristic imprint of protein binding. In a preferred embodiment this process is aided by consensus sequences.

Disclosure of the Invention

The invention is directed to a method to determine oligonucleotide sequences that specifically bind proteins or other targets. The method is especially applicable to DNAs wherein a consensus sequence site is known. In this case, knowledge of the nature of the protein or other target which is bound is not necessarily a requisite. This technique has been applied to describe the nucleotide sequences responsible for binding certain basic helix-loop-helix (bHLH) proteins which are important in differentiation, specifically MyoD, cMYC and a previously undescribed protein from reticulocytes.

Accordingly, in one aspect, the invention is directed to a method to determine an oligonucleotide sequence which binds specifically to a target ligand, which method comprises providing a mixture containing oligomers having portions which form a random set of sequences and portions which permit amplification of the

-6-

oligomers, treating the oligomer mixture with the target substance to form complexes between the target and the oligonucleotides bound specifically thereto, separating the complexes from the unbound members of the oligonucleotide mixture, recovering the complexed oligonucleotide(s) and amplifying these. This process will generally be repeated over several rounds of complexation, separation and amplification. When a mixture of sufficient binding affinity is obtained, this is followed by sequencing the recovered and amplified oligonucleotide(s) which had been complexed with the target. In a preferred embodiment, the mixture of oligonucleotides having random sequences also contains a consensus sequence known to bind the target.

In other aspects, the invention is directed to oligonucleotides identified by the above method, and to oligonucleotide sequences which bind specifically to MyoD, cMYC, and a bHLH protein from reticulocytes. In still another aspect, the invention is directed to complexes comprising target substance and specifically bound oligomer in a cell-free environment.

In still other aspects, the invention is directed to oligomers which contain sequences that bind specifically to target substances, and to the use of these oligomers in therapy, diagnostics, and purification procedures.

Brief Description of the Drawings

Figure 1 shows a diagrammatic representation of the method of the invention.

Figure 2 shows the DNA sequences of four oligomers used to illustrate the method of the invention.

Figure 3 shows typical separation results on an electrophoretic mobility shift assay (EMSA) of free

35

-7-

oligonucleotides and bound oligonucleotides to a MyoD-containing fusion protein.

Figure 4 shows typical sequencing results obtained from a control and complexed oligonucleotide recovered from the gel of Figure 3.

Figure 5 shows the electrophoretic mobility separation (EMSA) of complexes formed by proteins obtained by in vitro transcription/translation with random and nonrandom oligonucleotide probes.

Figure 6 is a higher exposure of the EMSA results of Figure 5, along with a comparable exposure of an EMSA obtained from an additional complexation reaction.

Figure 7 shows the results of EMSA separations of oligomers retrieved by the process of the invention after additional rounds of complexation, separation, amplification and recovery.

Figure 8 shows sequencing results of the control oligonucleotide mixture and various selected oligomers from the mixture obtained from the complexes shown in Figure 7.

Figure 9 is a summary of the sequences of oligonucleotides obtained by the selection process of the invention.

Figure 10 shows an EMSA separation of proteins from myoblast and MEL cell extracts complexed to oligomer selected using binding to a crude reticulocyte lysate.

Figure 11 shows EMSA results of recoveries after selection by method of the invention from randomized oligomers using cMYC fusion proteins.

Figure 12 shows the results of sequencing the recovered oligomers of Figure 11 after three rounds of selection.

Modes of Carrying Out the Invention

The invention is directed to a method which permits the recovery and deduction of oligomeric sequences that bind specifically to desired targets, including proteins. Therefore, as a result of application of this method, oligonucleotides which contain the specifically binding sequences can be prepared and used in oligonucleotide-based therapy and in other applications.

For example, these oligonucleotides can be used as a separation tool for retrieving the substances to which they specifically bind. By coupling the oligonucleotides containing the specifically binding sequences to a solid support, for example, proteins or other cellular components to which they bind can be recovered in useful quantities. In addition, these oligonucleotides can be used in diagnosis by employing them in specific binding assays for the target substances. When suitably labeled using detectable moieties such as radioisotopes, the specifically binding oligonucleotides can also be used for in vivo imaging or histological analysis.

"Oligomers" or "oligonucleotides" includes RNA or DNA sequences of more than one nucleotide in either single chain or duplex form and specifically includes short sequences such as dimers and trimers, in either single chain or duplex form, which may be intermediates in the production of the specifically binding oligonucleotides.

As used herein, "specifically binding oligonucleotides" refers to oligonucleotides which are capable of forming complexes with an intended target substance in an environment wherein other substances in the same environment are not complexed to the oligonucleotide. In general, a minimum of approximately

-9-

10 nucleotides, preferably 15 nucleotides, are necessary to effect specific binding. The only apparent limitations on the binding specificity of the target/oligonucleotide couples of the invention concern
5 sufficient sequence to be distinctive in the binding oligonucleotide and sufficient binding capacity of the target substance to obtain the necessary interaction. Oligonucleotides of sequences shorter than 10 may also be
10 feasible if the appropriate interaction can be obtained in the context of the environment in which the target is placed. Thus, if there are few interferences by other materials, less specificity and less strength of binding may be required.

As further explained below, the specifically
15 binding oligonucleotides need to contain the sequence-conferring specificity, but may be extended with flanking regions and otherwise derivatized.

After application of the method of the invention has resulted in the identification of one or
20 more oligonucleotides that bind specifically to target, the specifically binding oligonucleotides may be sequenced, and then resynthesized in any convenient form for the intended use. As an oligonucleotide having the identified sequence or a deliberately modified form
25 thereof can be synthesized de novo on the basis of this information, the oligonucleotides identified by the method of the invention in effect can include modifications both to the backbone structure and to the bases substituted thereon that may confer desirable
30 properties, such as enhanced permeation or increased stability with respect to nucleases. In general, the information obtained by analysis of the oligonucleotide pool obtained as a result of the invention process is thus used in synthesis of oligonucleotides with any
35 desired modification.

-10-

Thus, the oligonucleotides that comprise the sequences specifically binding to target substance may be conventional DNA or RNA moieties, or may be "modified" oligomers which are those conventionally recognized in the art. As the oligomers of the invention are defined also to include intermediates in their synthesis, any of the hydroxyl groups ordinarily present may be protected by a standard protecting group, or activated to prepare additional linkages to additional nucleotides, or may be conjugated to solid supports. The 5' or 3' terminal OH is conventionally activated; the alternate terminal 3' or 5' OH may be protected. In the oligonucleotide products and intermediates, one or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein P(O)O of the conventional phosphodiester is replaced by P(O)S, P(O)NR₂, P(O)R, P(S)S, P(O)OR', CO, or CNR₂, wherein R is H or alkyl (1-6C) and R' is alkyl (1-6C); in addition, this group may be attached to adjacent nucleotide through O or S. Not all linkages in the same oligomer need to be identical.

While ordinarily the randomized portions of the oligonucleotides described below will contain the conventional bases adenine, guanine, cytosine, and thymine or uridine, included within the invention are oligonucleotides that which incorporate analogous forms of purines and pyrimidines.

"Analogous" forms of purines and pyrimidines are those generally known in the art, many of which are used as chemotherapeutic agents. An exemplary but not exhaustive list includes 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, dihydrouracil, inosine, N6-iso-

-11-

pentenyladenine, 1-methyladenine, 1-methylpseudouracil,
1-methylguanine, 1-methylinosine, 2,2-dimethylguanine,
2-methyladenine, 2-methylguanine, 3-methylcytosine,
5-methylcytosine, N6-methyladenine, 7-methylguanine,
5 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiour-
acil, beta-D-mannosylqueosine, 5'methoxycarbonylmethyl-
uracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyl-
adenine, uracil-5-oxyacetic acid methylester,
uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil,
10 queosine, 2-thiocytosine, 5-methyl-2-thiouracil,
2-thiouracil, 4-thiouracil, 5-methyluracil,
N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic
acid (v), pseudouracil, queosine, 2-thiocytosine, and
2,6-diaminopurine.

15 In most instances, the conventional bases will
be used in applying the method of the invention; substi-
tution of analogous forms of purines and pyrimidines may
be advantageous in designing the final product.

The oligonucleotides containing the specific
20 binding sequences discerned through the method of the
invention can also be derivatized in various ways. For
example, if the oligonucleotide containing the
specifically binding sequence is to be used for separ-
ation of the target substance, conventionally the oligo-
25 nucleotide will be derivatized to a solid support to
permit chromatographic separation. If the oligonucleo-
tide is to be used to label cellular components or other-
wise for attaching a detectable moiety to target, the
oligonucleotide will be derivatized to include a radio-
30 nuclide, a fluorescent molecule, a chromophore or the
like. If the oligonucleotide is to be used in specific
binding assays, coupling to solid support or detectable
label, and the like are also desirable. If to be used in
therapy, the oligonucleotide may be derivatized to
35 include ligands which permit easier transit of cellular

-12-

barriers, toxic moieties which aid in the therapeutic effect, or enzymatic activities which perform desired functions at the targeted site. The oligonucleotide may also be included in a suitable expression system to provide for in situ generation of the desired sequence.

In general, the oligonucleotides identified according to the method of the invention, and, if desired, synthesized de novo either in native or modified form are useful in a manner analogous to antibodies or specifically immunoreactive fragments thereof. These invention oligonucleotides are characterized by their ability specifically to bind the intended target molecule in both simple and complex environments. Thus, the formation of an oligonucleotide-target complex may be formatted in procedures analogous to those employed in immunoassay procedures. A wide range of such protocols is known in the art, and includes both direct and competitive formats, and involves employment of a wide range of detection techniques. Similarly, as antibodies may be used in diagnostic and therapeutic applications, as well as in the control of cell growth and differentiation, so too may the oligonucleotides of the invention.

25 The Invention Method of Oligonucleotide Identification

The oligonucleotides used as starting materials in the process of the invention to determine specific binding sequences may be single-stranded or double-stranded DNA or may be RNA. Double-stranded DNA is preferred. In any case, the starting material oligonucleotide will contain a randomized sequence portion flanked by primer sequences which permit the application of the polymerase chain reaction to the recovered oligonucleotide from the complex. These flanking sequences may also contain other convenient features such as

-13-

restriction sites which permit the cloning of the amplified sequence.

The randomized portion may be constructed using conventional solid phase techniques using mixtures of nucleotides at the positions where randomization is desired. Of course, any degree of randomization may be employed; some positions may be randomized by mixtures of only two or three bases rather than the conventional four; randomized positions may alternate with those which have been specified. Indeed, it is helpful if some portions of the candidate randomized sequence are in fact known. In the illustration set forth in the examples below, the target substances are proteins for which consensus sequences are known.

While the method of the invention is illustrated using proteins as target substances, any ligand which is of sufficient size to be specifically recognized by an oligonucleotide sequence can be used as the target. Thus, glycoproteins, proteins, carbohydrates, membrane structures, receptors, lipids, organelles, and the like can be used as the complexation targets. As illustrated below, however, the process is greatly aided if a consensus sequence for the target is known. A particular illustration of this application is set forth in the examples below with respect to the basic-HLH domains which characterize a number of proteins involved in development and differentiation of tissues. These proteins include a region of basic amino acids which are followed in sequence (N-C) by a helix-loop-helix region which is thought to mediate multimerization of the proteins. The multimerization results in positioning the basic regions so as to make specific contacts with the DNA.

It is already known that DNA sequences which bind proteins containing bHLH regions contain a

-14-

palindromic consensus region CANNTG. Proteins containing the bHLH region are produced by the gene E2A, MyoD (which is associated with myogenesis and expression of muscle-specific genes), cMYC (an oncogene), and other genes involved in development described below. The presence of the consensus sequence and the availability of the corresponding proteins is helpful in applying the method; however, the method can be applied even where there is no consensus sequence, if the target is available. The method can also be applied to retrieve unknown proteins, especially where a consensus sequence is known.

An outline of the procedure of the invention is shown in Figure 1. The steps of this process result in "Selected and Amplified Binding-Sites" (SaABs). As illustrated, a mixture of oligonucleotides is synthesized with random sequences in the intended binding site that are flanked by suitable regions for hybridization to primers for use in PCR. As shown in Figure 1, item 1, a single strand DNA is prepared with random nucleotide sequence NNN where the region for primer hybridization, A, is shown at the 3' end. The oligomer is formed into a duplex by synthesizing the opposite strand, which now has primer hybridization regions A and B. This is incubated with the target, in this case a protein, and the complexes shown as item 3 are separated from the uncomplexed duplexes using the mobility shift in electrophoresis (EMSA). The bound templates are rescued by PCR and amplified for sequencing. The original double-stranded oligonucleotide in item 2 is also amplified as a control. The resulting amplified sequences are applied to sequencing gels to determine the nature of the "ABC" counterparts of the random nucleotides selected. The entire process is repeated using the recovered and amplified duplex until sufficient resolution is obtained.

35

-15-

The procedure shown in Figure 1 is merely illustrative. The mixture of oligonucleotides may be comprised of single-stranded DNA or RNA as well as the double-stranded DNA shown. In this instance, the primer sequences flank the randomized portion on a single oligonucleotide chain. The separation of the portion of the mixture which binds to target substance may be conducted in any convenient manner. For example, rather than relying on a difference in electrophoretic mobility of the complex as compared to the unbound oligonucleotide, the target substance may be coupled with a solid support and the oligonucleotide mixture applied to the support. The portion of the mixture which fails to bind to the coupled target is then simply washed from the support, leaving behind the complexed portion of the mixture. Thus, in general, the procedure simply involves complexation of the mixture with the target, separation of the complexed oligonucleotides from those failing to participate in the complex and rescue of the complexed oligonucleotides by amplification. The amplification of the oligonucleotides which bind to the target may be conducted either while the complex is still intact or after prior separation of the complexed oligonucleotides from the target.

In general, more than one "round" of binding, separation of the complex, and amplification will be required in order to achieve a set of appropriately binding oligonucleotides. The process is simply repeated using the recovered subset of binding nucleotides as starting material in subsequent rounds until a mixture containing sufficient binding affinity is obtained. In general, it will be desirable to sequence this specifically binding subset to determine consensus sequences in the specifically binding oligomers. As set forth above, the members of this subset may then be

-16-

synthesized de novo, thus permitting the preparation of oligonucleotides that contain either base modifications, backbone modifications, or both.

5 Utility of the Retrieved Sequence

Accordingly, the oligomers of the invention which contain specifically binding nucleotide sequences are useful in therapeutic, diagnostic and research contexts. In therapeutic applications, the oligomers are
10 utilized in a manner appropriate for oligonucleotide therapy in general--as described above, oligonucleotide therapy as used herein includes any use of oligonucleotides as medicaments, whether this involves targeting a specific DNA or RNA or targeting any other substance
15 through complementarity or through any other specific binding means, for example, sequence-specific orientation in the major groove of the DNA double-helix, or any other specific binding mode. For such therapy, the oligomers of the invention can be formulated for a variety of modes
20 of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, PA, latest edition.

25 For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the oligomers of the invention are formulated in liquid solutions, preferably in physiologically compatible buffers
30 such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by
35 transmucosal or transdermal means, or the compounds can

-17-

be administered orally. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays, for example, or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics.

For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams, as is generally known in the art.

The oligonucleotides may also be employed in expression systems, which are administered according to techniques applicable, for instance, in applying gene therapy.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target substances to which they specifically bind. Such diagnostic tests are conducted by contacting a sample with the specifically binding oligonucleotide to obtain a complex which is then detected by conventional means. For example, the oligomers may be labeled using radioactive, fluorescent, or chromogenic labels and the presence of label bound to solid support to which the target substance has been bound through a specific or nonspecific binding means detected. Alternatively, the specifically binding oligomers may be used to effect initial complexation to the support. Means for conducting assays using such oligomers as specific

35

-18-

binding partners are generally known to track those for standard specific binding partner based assays.

It may be commented that the mechanism by which the specifically binding oligomers of the invention interfere with or inhibit the activity of a target substance is not always established, and is not a part of the invention. The oligomers of the invention are characterized by their ability to target specific substances regardless of the mechanisms of targeting or the mechanism of the effect thereof.

For use in research, the specifically binding oligonucleotides of the invention are especially helpful in effecting the isolation and purification of substances to which they bind. For this application, typically, the oligonucleotide containing the specific binding sequences is conjugated to a solid support and used as an affinity ligand in chromatographic separation of the target substance. The affinity ligand can also be used to recover previously unknown substances from sources which do not contain the target substance by virtue of binding similarity between the intended target and the unknown proteins. Furthermore, as data accumulate with respect to the nature of the nonoligonucleotide/oligonucleotide-specific binding, insight may be gained as to the mechanisms for control of gene expression.

The following examples are meant to illustrate, but not to limit the invention.

Example 1

DNA Binding MyoD Target Proteins

The oligonucleotide sequences in the randomized mixtures were synthesized using standard solid-phase synthesis techniques and are shown in Figure 2. As shown in Figure 2, the MCK (muscle creatine kinase enhancer) is a naturally occurring sequence known to bind MyoD.

-19-

Oligomers D1, D2 and D3 have various locations of randomization of sequence, and further contain regions for coupling to PCR primers shown as B and A' at the 5' and 3' ends, respectively. Primer A is

5' -TCCGAATTCCTACAG-3'

and primer B is

5' -AGACGGATCCATTGCA-3'.

These contain restriction enzyme sites for convenience. The double-stranded D1-D3 templates were generated by annealing the oligonucleotide to a 10-fold molar excess of primer A, synthesizing the complementary strand using Klenow fragment of *E. coli* DNA polymerase and purifying the template on a 12% polyacrylamide gel. The templates were end-labeled using the kinase reaction of Davis, R.L., et al., *Cell* (1990) 60:733. The MCK double-stranded template was obtained from a kinased oligonucleotide annealed to its complement.

As shown in Figure 2, in D1 and in D2, the randomization obliterates a portion of the consensus sequence in each case. In D3, randomization is limited to two nucleotides upstream, two nucleotides downstream, and the two nucleotides between the members of the consensus motif.

Complexation was conducted using approximately 200 ng of glutathione-MyoD bacterially produced fusion protein and either 0.15 ng of the MCK template or 0.30 ng of the random sequence templates (about 6×10^4 cpm each) as described by Lassar, A.B., et al., *Cell* (1989) 58:823 but using 100 ng of poly(dI-dC) in each incubation. EMSA was performed on a 6% polyacrylamide gel as described in Davis, R.L., et al., *Cell* (1990) 60:733.

The results of the incubation of glu-MyoD with MCK, D1 and D2, subjected to EMSA, are shown in Figure 3. As indicated in the figure, the fusion protein binds readily to the MCK sequences and less well to D1 and D2,

-20-

as large numbers of the oligomers in the randomized mixture have inappropriate sequences.

To reisolate the complexed templates, a slice approximately 0.3 cm wide was excised from the dried-down gel including the 3 MM (Whatman) paper backing. The gel slices were incubated at 37°C overnight in 0.5 ml of 0.5 M ammonium acetate, 10 mM MgCl₂, 1 mM EDTA, and 0.1% SDS. Approximately 50% of the radioactivity was recovered. After addition of 5 µg of tRNA carrier, the eluate was extracted twice each with phenol and with chloroform:isoamyl alcohol, 24:1, and precipitated with ethanol. The precipitates were brought to 0.3 M sodium acetate and reprecipitated with ethanol.

About 1/5 of the resuspended sample was amplified for 35 cycles of PCR in 100 µl reaction using primers A and B, under the standard conditions described by Saiki, R.K. in PCR Technology, A.J. Ehrlich ed. (Stockton Press, NY) 1989, pages 7-16, following optimization of Mg⁺² concentration. Under carefully controlled conditions, a test reaction that contained 1 pg of starting template yielded approximately 100 ng of product. Reactions performed on the material excised from EMSA yielded 30-100 ng DNA. The products of the reaction were purified on 14% polyacrylamide gels and eluted and purified as set forth above.

The recovered and amplified complexed oligomers were then sequenced using labeled primer A or B and the termination step of the Sequenase procedure marketed by United States Biochemical Co. as follows. The primers were labeled using a kinase reaction to 1-2x10⁶ cpm/ng and unincorporated label was removed using a Sephadex G50 spin column. 10 ng labeled primer were mixed with about 5 ng purified oligomer to be sequenced in a 12 µl reaction that contained 1 µl Sequenase Mn⁺² buffer and 2 µl 5 x Sequenase buffer. The reaction was incubated at

-21-

95°C for 5 min and then quick spun at room temperature for 1 min. The reaction was placed on ice and to it were added 1 μ l 0.1 M dithiothreitol and 2 μ l of diluted Sequenase 2.0 enzyme (1:8 in ice-cold TE, pH 7.4). 3.5 μ l of this mixture were added to 2.5 μ l of each of the Sequenase dGTP termination mixes and incubated at 45°C for 4 min. The reactions were terminated by adding 4 μ l Sequenase stop solution. (Mn^{+2} buffer was omitted from reactions performed with dTTP termination mixes.)

The reactions were run on a 14% denaturing polyacrylamide sequencing gel containing 8 M urea in TBE. 1.5 μ l of a reaction were loaded into each well with the exception of the "C" reaction in sequences generated with primer B as the nonrandom bases appearing in the C lane were generally fainter than those in the corresponding G, A, and T lanes. This difference was compensated by loading 2.5 μ l of the C reaction. Before fixing the gel in 10% acetic acid and 10% methanol, the large excess of unreacted primer was cut away to prevent its diffusion.

The results of sequencing the bands of Figure 3 are shown in Figure 4. As shown in Figure 4, preferential recovery of the consensus sequence embodiments from randomized portions of the consensus sequence was obtained. There was also a preference for thymidine in position 4 (see Figure 2), different from the cytosine present in the MCK sequence.

Example 2

DNA Sequences Targeting Various Proteins

As the procedure in Example 1 established the criticality of the consensus sequence in binding to MyoD, the D3 oligomer which contains this sequence was used in subsequent studies. While D3 contains the consensus sequence, it is randomized in the immediate proximity. Various proteins which are associated with

-22-

differentiation, including MyoD, E2A, E12, and E47, were synthesized by in vitro translation from DNAs, some of which are reported by Murre, C., et al., Cell (1989) 56:777. The transcribed sequences were prepared from a mouse MyoD cDNA, a human E12 cDNA (E12R) and a human E47 cDNA as described by Benezra, R., et al., Cell (1990) 61:49. About 2.5 μ l of a 50 μ l reticulocyte lysate (Promega) in vitro translation reaction were then used to test binding with the randomized oligomers. Homodimers, homomultimers, heterodimers and heteromultimers were formed from these protein products. To form heteromultimers, separate translation reactions were mixed prior to DNA binding and incubated at 37°C for 20 min before adding to a binding reaction cocktail. The protein preparations were then incubated with either D2 or D3 as follows.

The final binding reaction to test randomized oligomers contained 20 mM Hepes, pH 7.6, 50 mM KCl, 1 mM dithiothreitol, 1 mM EDTA, 8% glycerol, 0.1 μ g polydI/dc and 2 μ g of a 50 bp single-stranded oligonucleotide, both added as nonspecific competitors.

Each binding reaction contained the in vitro synthesized protein species at about 6.9×10^{-11} M and either 0.15 ng of MCK or 0.30 ng of D2 or D3 labeled templates providing a protein:DNA molar ratio of about 0.18. Binding reactions were performed at room temperature for 20 min and immediately subjected to EMSA.

The results of application of these incubation mixtures to EMSA are shown in Figure 5. These results indicate that MCK binds strongly to E12/MyoD, E47 and E47/MyoD; D3 binds only to E47. However, a longer exposure of these gels (Figure 6), along with a gel run on analogous reaction mixtures using oligomer D2 shows complexation of MCK with all of the tested samples and of D3 with E12/MyoD, and E47/MyoD in addition to E47.

-23-

Bands that were excised from the gels shown in Figure 6 were subjected to three additional rounds of incubation, EMSA, and PCR amplification. In such subsequent rounds, about 5 ng of the purified amplified template were labeled for one cycle in a 20 μ l reaction containing 30 μ Ci of 32 P dTP, 50 mM each of dATP, dGTP and dCTP and 100 ng each of primers A and B in the standard PCR reaction buffer. The large excess of primers was added to insure that synthesis occurred on all templates in the reaction. Unincorporated label was removed over a 1 ml G50 spin column, and the reaction products were ascertained as being full-length. The binding reaction and EMSA were performed as above but with about 0.1 ng of the PCR-labeled template pool providing a protein:DNA molar ratio of about 0.54.

Because successive rounds enrich in the binding species, additional complexation was found. As shown in Figure 7, complexation putatively yielding sequence specificity in comparison to the controls was found between D3 and target proteins MyoD, E12/MyoD, E47, and E47/MyoD. D2 complexed with E12/MyoD, E47, and E47/MyoD. Importantly, Figure 7 further shows that reticulocyte factors other than the target sequences are also bound by selected D3 oligomers, particularly those selected by E12, E12/MyoD and E47/MyoD.

Figure 8 shows the results of sequencing performed as described above with some of the complexes shown in Figure 7 which were excised from the gel.

Figure 8A gives the results for the D3 control mixture showing complete heterogeneity at the six positions which were randomized. DNA sequences which had been selected by the invention process (Figures 8B-8F) showed positional preferences. For example, the D3 oligomer selected by complexation to MyoD (Figure 8B) showed a clear preference for T in positions 5 and 4,

-24-

retained some heterogeneity in positions 1 and -1, and showed some preference for A in positions -4 and -5. The reticulocyte lysate shown in Figure 8F apparently recognizes the sequence (G/A)CCAGTTG(N)A.

5 A summary of the results of the binding and sequencing experiments illustrated in Figure 8 is shown in Figure 9. The preset position choices are shown on shaded backing, the assignment preferences that are absolute or nearly so are indicated with capital letters, 10 and incomplete preferences are printed in lower case. However, a bar over the letter indicates exactly the opposite--the base is never found at the indicated position (capitals) or only weakly represented (lower case).

15

Example 3

Use of the Lysate D3 Template to Retrieve Specific Proteins

Nuclear extracts of P2 myoblasts (Lassar, A.B., 20 et al., Cell (1986) 47:649) and of a murine erythro-leukemia (MEL) nuclear cell extract were used as the source of target protein in the binding, EMSA, and amplification rounds set forth above. The P2 myoblast extract was prepared as described by Dignam, J.D., et 25 al., Nucleic Acids Res (1983) 11:1475, except that the extract was not dialyzed; the MEL cell extract was prepared as described by Gorski, K., et al., Cell (1986) 47:767. Both MCK and the lysate-derived D3 template were used in complexation reactions, conducted as described 30 above under the following conditions: P2 myoblast binding reactions contain 20 mM Hepes, pH 7.6, 1.5 mM MgCl₂, 50 mM NaCl, 5% glycerol and 500 ng poly(dI/dC). MEL cell binding reactions were conducted as described in Example 1, except that each reaction contained 2 µg of 35 polydI/dC. Both binding reactions were incubated at room

-25-

temperature for 20 minutes and then immediately subjected to EMSA on 5% polyacrylamide gels at 200 V at 4°C. The results are shown in Figure 10. As shown, the lysate-selected D3 binds to factors in the MEL extract; and to different factors from that bound by MCK in the myoblast cell extract. These previously unidentified target proteins are therefore recoverable by virtue of their ability to bind lysate-selected D3.

Example 4

DNA Sequences Specific for cMYC Protein

A bacterially produced glutathione S-transferase (GST) fusion protein which contains the C-terminal 92 amino acids of human cMYC (cMYC-C92) was used as the target protein. This fusion protein includes the bHLH domain and leucine zipper. The DNA template used was D6 as shown in Figure 2 which has random sequences flanking the consensus sequence and A and B primers as set forth above. Several rounds of complexation, EMSA, and amplification were required to recover the preferred DNA binding sequences as shown in Figure 11. Figures 11, lanes 2 and 3, indicate the results from the second and third rounds of the complexation/separation/amplification cycle.

Figure 12 shows the sequencing results. Amplified D3 was used as a control. As indicated in the figure, the two bases internal to the consensus sequence have been identified, but heterogeneity in the flanking sequences persists.

30

35

-26-

Claims

1. A method to identify an oligonucleotide sequence which specifically binds a target substance, which process comprises:
- 5 incubating said target substance with a mixture of randomized oligonucleotide sequences under conditions wherein complexation occurs with some, but not all, members of said mixture;
- 10 separating complexed from uncomplexed oligonucleotides;
- recovering and amplifying the complexed oligonucleotide(s) from said complexes; and
- 15 optionally determining the sequence of the recovered oligonucleotide(s).
2. The method of claim 1 wherein said target substance is known to bind to a consensus sequence, and the oligonucleotide mixture is not randomized at the
- 20 positions of the consensus sequence.
3. The method of claim 1 wherein the oligonucleotide mixture is double-stranded DNA.
- 25 4. The method of claim 3 wherein the target substance is a protein and contains a basic loop-helix-loop (bHLH) region and the consensus region is CANNTG.
- 30 5. The method of claim 1 wherein said separating is conducted using electrophoresis mobility shift assay (EMSA) and wherein said recovering and amplifying is through the conduct of the polymerase chain reaction (PCR).
- 35

-27-

6. A mixture of oligonucleotide segments useful as a starting material in the recovery of an oligonucleotide sequence specifically binding to a target substance, which mixture comprises a randomized set of nucleotide sequences wherein in each member of the set said segment containing a randomized DNA sequence is flanked with primer sequences for PCR.

7. The mixture of claim 8 wherein said segments are double-stranded DNA segments.

8. An oligonucleotide comprising a sequence specific for a target substance, in purified and isolated form, identified by the process of claim 1.

9. The oligonucleotide of claim 8 wherein the target substance is MyoD or cMYC.

10. A method for recovering a target substance from a sample, which method comprises contacting said sample with at least one oligonucleotide containing a sequence specifically binding for said substance wherein said oligonucleotide has been identified by the process of claim 1 under conditions wherein said substance and said oligonucleotide form a complex;

separating the complex from other materials in the sample; and

recovering the substance from the complex.

11. A method to modify target cells or tissues in a subject, which method comprises administering to a subject in need of such modification an oligonucleotide which comprises a sequence specifically binding to a substance characteristic of said target cell or tissue, wherein said oligonucleotide is optionally derivatized

-28-

with a moiety which enhances said modification and wherein said oligonucleotide has been identified by the process of claim 1.

5 12. A method to determine the presence or absence of an analyte in a sample, which method comprises treating said sample with an oligonucleotide containing a sequence which specifically binds said analyte under conditions wherein a complex between said oligonucleotide
10 and analyte is formed when analyte is present, and detecting the presence or absence of the complex,
 wherein said oligonucleotide has been identified by the process of claim 1.

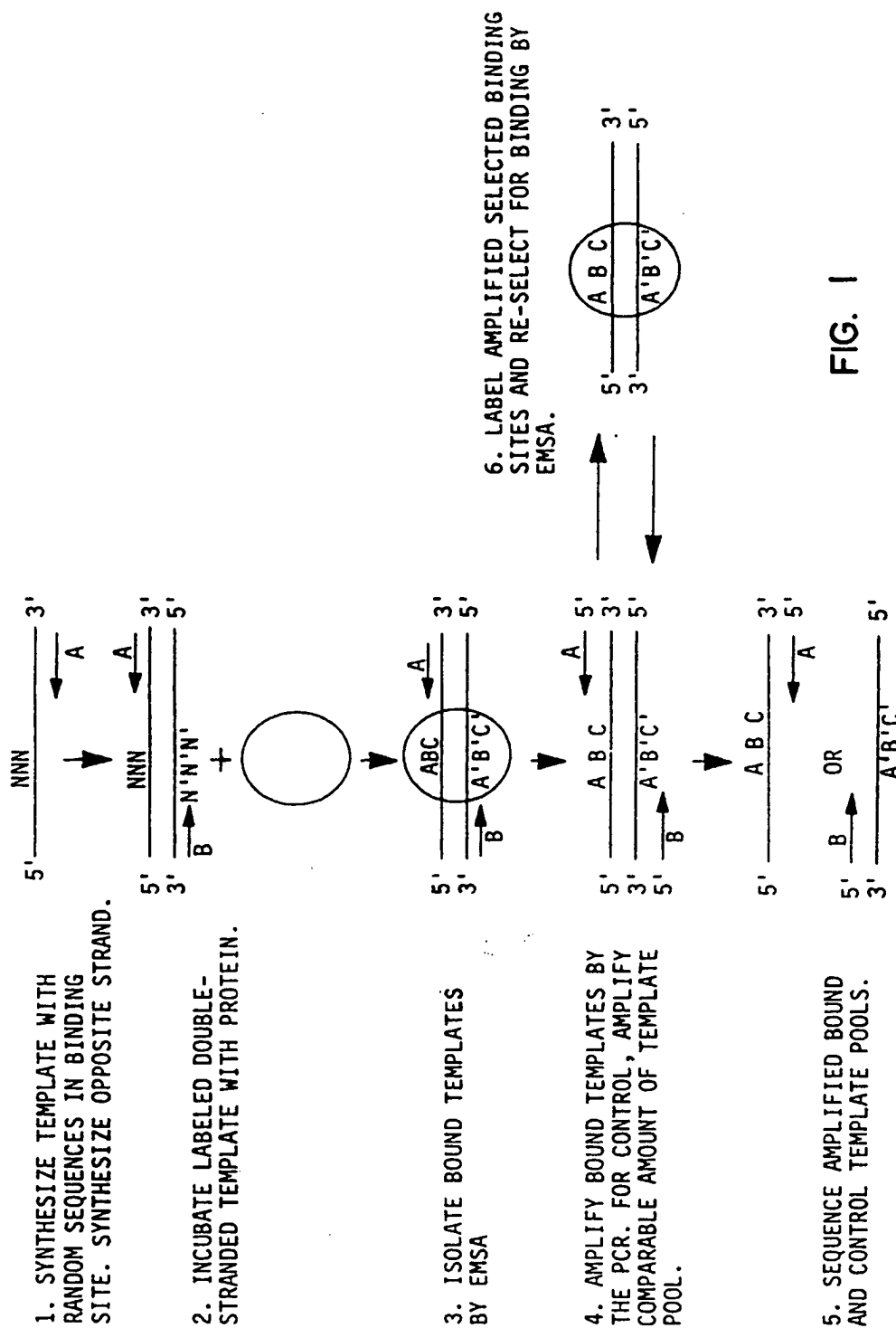
15 13. A complex which comprises a target substance and at least one specifically bound oligonucleotide wherein said oligonucleotide has been identified by the process of claim 1, which complex is
20 free of cellular components.

25

30

35

1 / 12



SUBSTITUTE SHEET

2 / 12

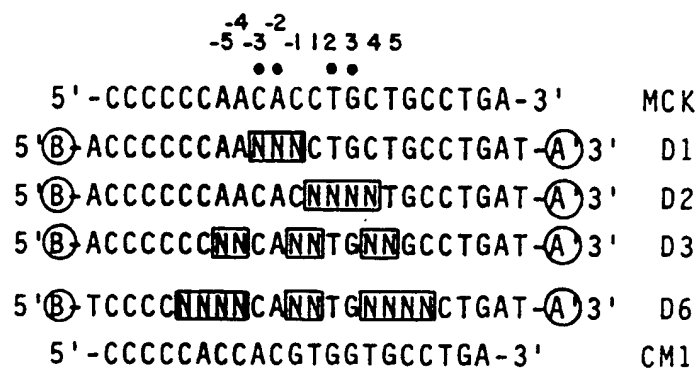


FIG. 2

SUBSTITUTE SHEET

3 / 12

-MCK
-D1
-D2



1 2 3
GluMyoD

FIG. 3

SUBSTITUTE SHEET

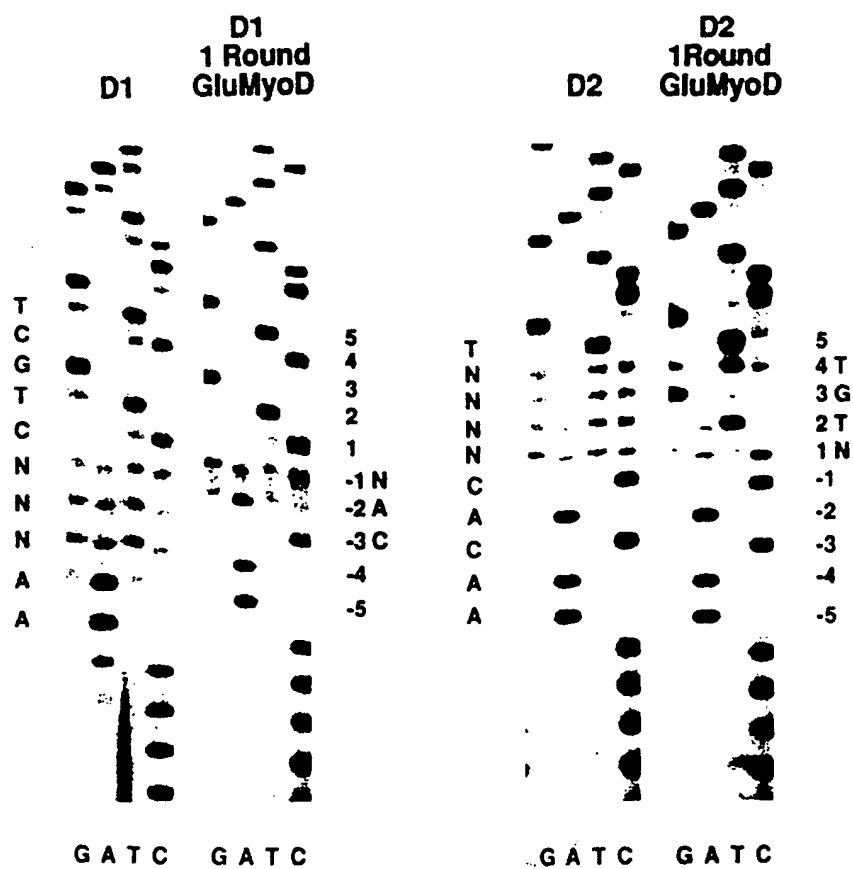


FIG. 4

SUBSTITUTE SHEET

5 / 12

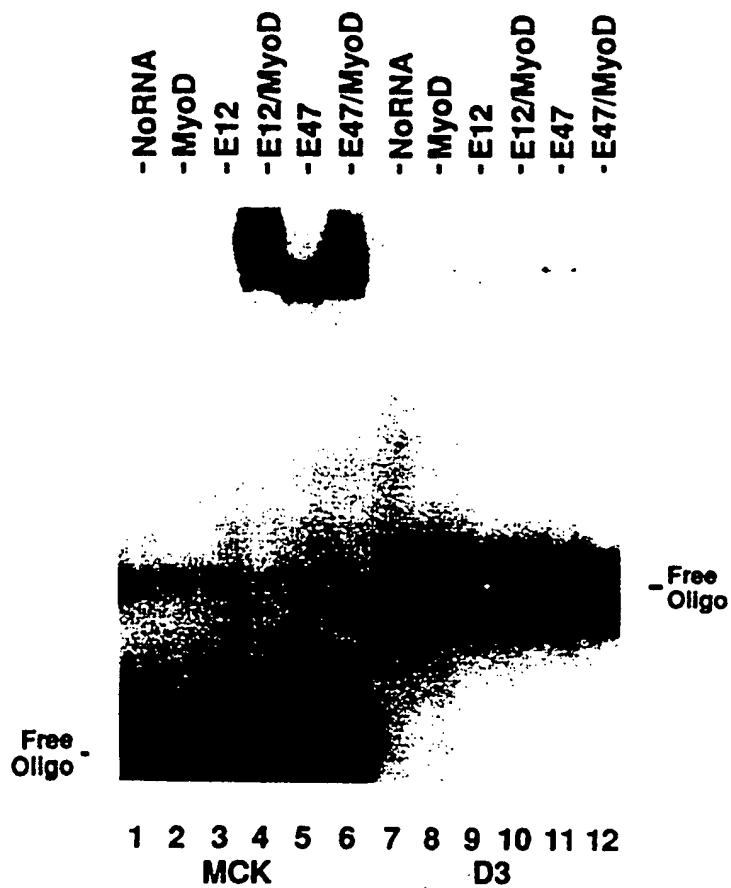


FIG. 5

SUBSTITUTE SHEET

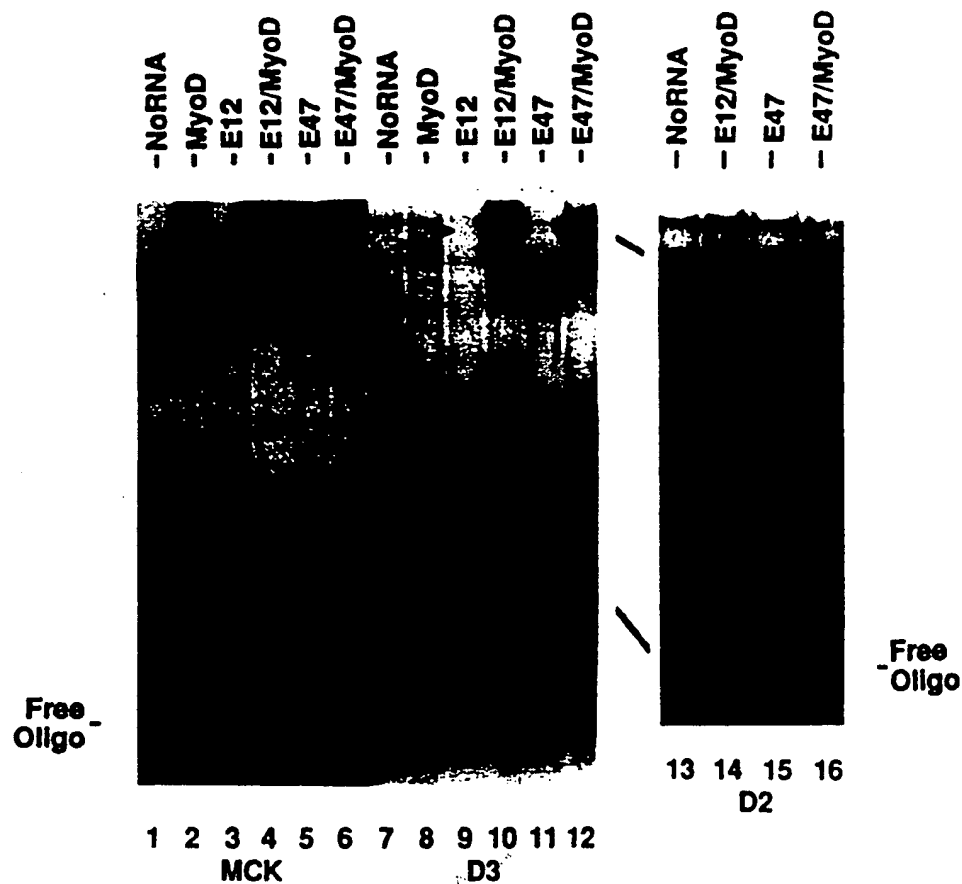


FIG. 6

SUBSTITUTE SHEET

7 / 12

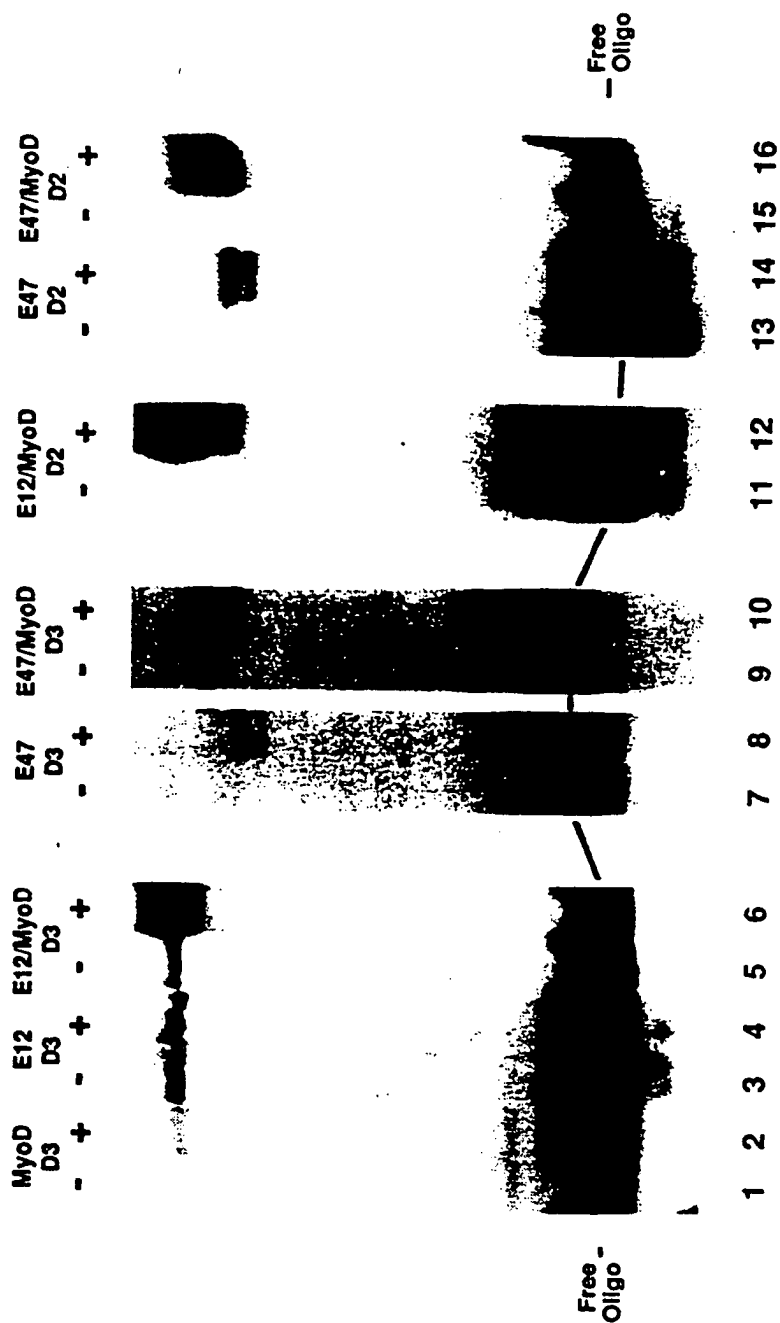


FIG. 7

SUBSTITUTE SHEET

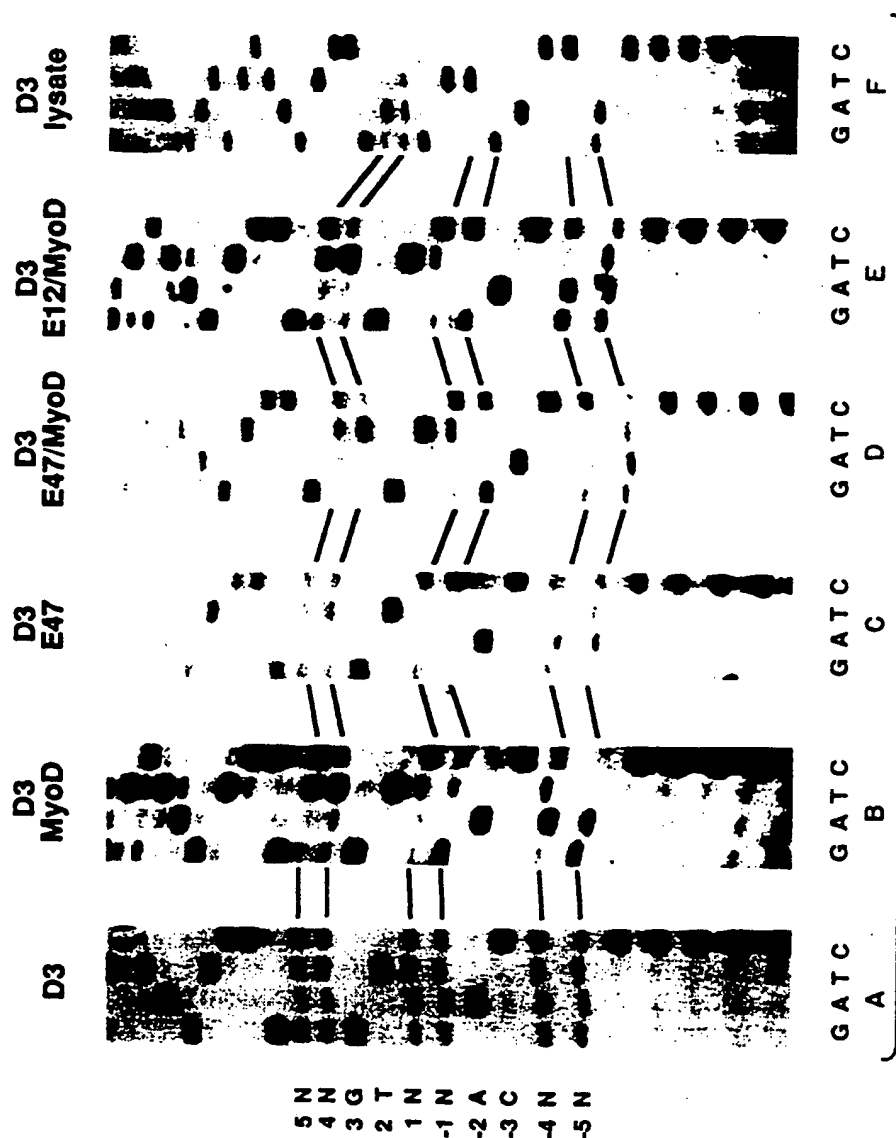


FIG. 8

SUBSTITUTE SHEET

9 / 12

A: PREFERENCES:		-5	-4	-3	-2	-1	1	2	3	4	5	
MyoD	D3	G	A	C	A	G	C	T	G	a	T	SYMMETRICAL
		A	t					T		T	C	
E47	D3	N	T	C	A	C	C	T	G	A	a	ASYMMETRICAL
						g	g					
	D2			C	A	C	C	T	G	A		
CONSENSUS		N	T	C	A	C	C	T	G	A	a	
						g	g					
LYSATE	D3	G	C	C	A	G	T	T	G	c	g	ASYMMETRICAL
		A								A		
E47/MyoD	D3	N	T	C	A	G	T	T	G	T	N	ASYMMETRICAL
						C	C					
	D2			C	A	C	C	T	G	T		
CONSENSUS		N	T	C	A	C	C	T	G	T	N	
E47/MyoD	D3	C	T	C	A	C	C	T	G	T	t	ASYMMETRICAL
						g	t				C	
				C	A	C	C	T	G	T		
CONSENSUS		C	T	C	A	C	C	T	G	T	t	
											C	
B: HALF-SITES:												
MyoD							C	T	G	T	T	
E47/E12		N	T	C	A	C					C	

FIG. 9

SUBSTITUTE SHEET

10 / 12

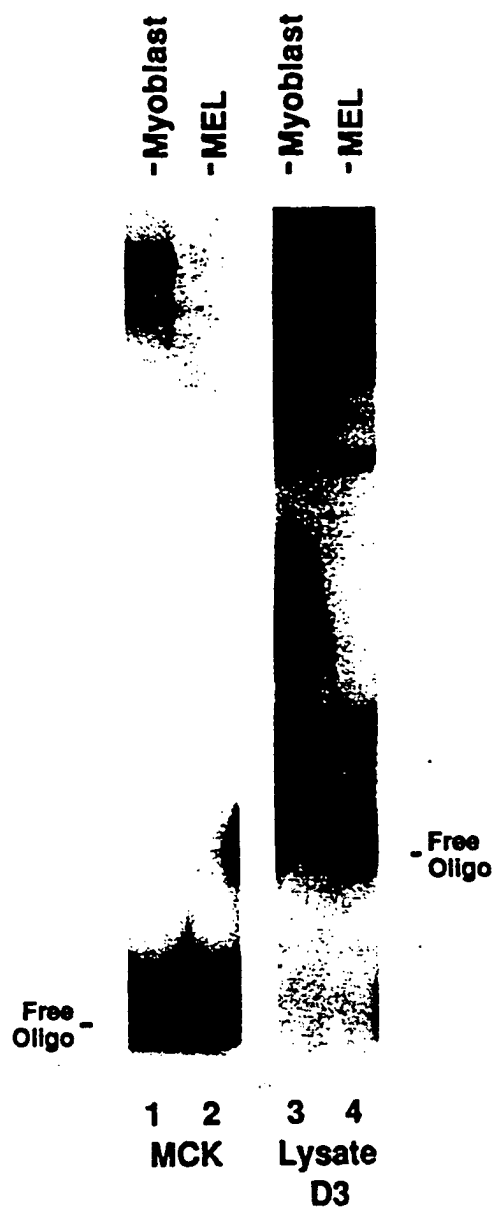


FIG. 10

SUBSTITUTE SHEET

11 / 12

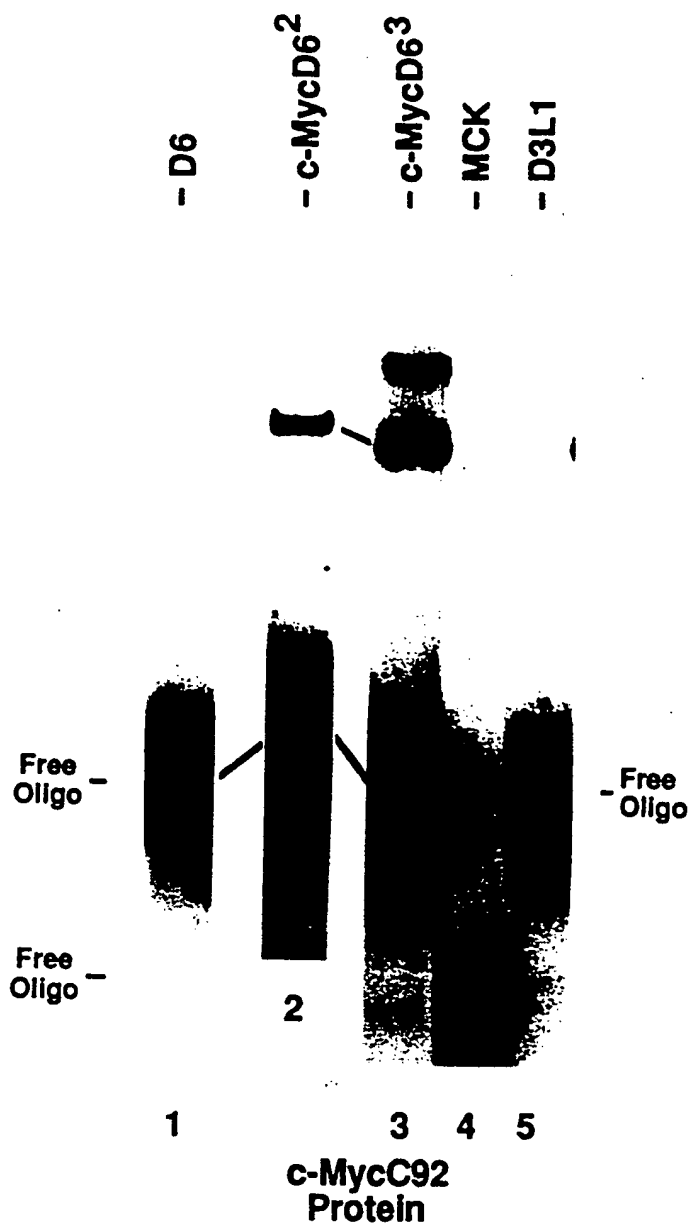


FIG. 11

SUBSTITUTE SHEET

12 / 12

D3 cMycD6³

7 N
6 N
5 N
4 N
3 G
2 T
1 N
-1 N
-2 A
-3 C
-4 N
-5 N
-6 N
-7 N



G
T
G
C
A
C

G A T C G A T C

FIG. 12

SUBSTITUTE SHEET

1p

0

0

6

5